DOCUMENT RESUME

ED 141 792 CS 003 580

AUTHOR Selden, Ramsay
TITLE Development of a Frequency-based Measure of Syntactic Difficulty for Estimating Readability.
PUB DATE May 77
NOTE 13p.; Paper presented at the Annual Meeting of the International Reading Association (22nd, Miami Beach, Florida, May 2-6, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Elementary Secondary Education; *Language Patterns; *Measurement Techniques; *Readability; Readability Formulas; Reading Materials; *Reading Research; Sentence Structure; *Syntax

ABSTRACT

Readability estimates are usually based on measures of word difficulty and measures of sentence difficulty. Word difficulty is measured in two ways: by the structural size and complexity of words or by reference to phonomena of language use, such as word-list frequency or the regularity of spelling patterns. Sentence difficulty is measured only in terms of size or complexity, despite models of the reading process which suggest that readers could use a knowledge of recurring syntactic patterns to economize their scan of the text. This study identified syntactic patterns which occurred in samples of student writing and text and counted how often each pattern recurred. McCall-Crabbs test passages were classified for the commonness of their sentence patterns according to these counts. The commonness of the patterns was significantly associated with the readability of the passages, with other factors controlled. The results bear implications for measuring readability and designing instruction. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED141792

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"Development of a Frequency-based Measure of Syntactic Difficulty for Estimating Readability"

Presented to the 1977 Annual Meeting of
International Reading Association
Miami Beach, Florida

Ramsay Selden
Department of Research Methodology
266 Ruffner Hall
University of Virginia
Charlottesville, Virginia 22903

(804) 924 - 3091
296 - 8282
296 - 1213

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
Ramsay Selden
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

CS 003 580

"Development of a Frequency-based Measure of Syntactic Difficulty for Estimating Readability"

Ramsay Selden
Department of Research Methodology
University of Virginia

Readability estimates are usually based on measures of word difficulty and measures of sentence difficulty. Word difficulty is measured in two ways: by the structural size and complexity of words or by reference to phenomena of language use such as word-list frequency or the regularity of spelling patterns. Sentence difficulty is measured only in terms of size or complexity, despite models of the reading process which suggest that readers could use a knowledge of recurring syntactic patterns to economize their scan of the text. This study identified syntactic patterns which occurred in samples of student writing and text, and counted how often each pattern recurred. McCall-Crabbs test passages were classified for the commonness of their sentence patterns according to these counts. The commonness of the patterns was significantly associated with the readability of the passages, with other factors controlled. The results bear implications for measuring readability and designing instruction.

"Development of a Frequency-based Measure of Syntactic Difficulty for Estimating Readability"

Ramsay Selden
Department of Research Methodology
University of Virginia

Purpose

The purpose of this study was to improve the methods used to measure the syntactic difficulty of text as a factor which affects readability. Specifically, the study was conducted in order to determine whether the syntactic patterns which are used in sentences are repeated with varying frequencies in the writing of school children or in typical textbook prose, and whether the frequency of this recurrence is associated with the readability of the patterns when they appear in text. If this were true, it would suggest whether readers use a sense of expectation of common sentence patterns to guide and economize their search for textual features which give cues to the meaning of the text. Further, if this variable were found to be useful, it would provide a basis for selecting the instructional materials presented to readers according to syntactic difficulty, and it could strengthen the basis for sequencing syntactically-based comprehension skills in an instructional program.

Conceptual Rationale

Readability formulae generally include two types of variable--measures of word difficulty and measures of sentence difficulty.

Lively and Pressey (1923) developed a readability estimation technique which was based on the placement of the words contained in a piece of text in Thorndike's frequency-stratified word-list. Later, Vogel and Washburne (1928), Gray and Leary (1935) and others developed techniques for estimating readability which included a measure of word familiarity based on a frequency-stratified word-list and a measure of sentence difficulty based on structural complexity or mean length. Widely used formulae such as those developed by Dale and Chall (1948), Spache (1953), and Dolch (1948) are based on measures of sentence length and on an index of the familiarity of the words determined by looking them up in a frequency-stratified word-list. Recent formulae developed by Fry (1968) and by Harris and Jacobson (1974) include different measures. The Fry method is based on mean sentence length and mean word length in syllables. The Harris-Jacobson method includes vocabulary familiarity, word length, sentence length, and the presence of words with irregular or uncommon sound-letter correspondences.

The variables which are generally used to measure word difficulty are of two types. The first type is based on the premise that the structural size and complexity of the word affects the difficulty with which it can be read; this type of measure is exemplified by measures of mean word length in syllables or letters, and by measures of the mean number of affixes present on words in the text. The second type of variable is based on the premise that phenomena of usage determine the familiarity of whole words, or

determine the ease with which a word can be decoded using sound-letter correspondances; this type is exemplified by word-list based measures of word familiarity, and by the Harris-Jacobson measure of sound-letter regularity.

While readability formulae include both types of word-difficulty variable, they include only measures of sentence difficulty which are based on the premise that structural size and complexity affect the readability of sentences. The frequency or regularity of sentence-structure phenomena has not been systematically identified and used as a factor affecting the readability of text. Bormuth (1969) studied a large number of sentence-structure characteristics in relation to readability, but was not able to conclude what general factors in sentence construction seemed to affect readability. Similarly, Granowski (1971) and Botel and Granowski (1972) report development of an index of sentence difficulty which includes consideration of structure familiarity, but the eclectic nature of the index precludes determination of the usefulness of this variable.

Models of the reading process suggest that readers can use expectations of sentence patterns to guide their search for cues to meaning. Smith (1971) posits a model of the reading process in which the reader uses an active, anticipatory approach to the text, along with the redundancy in the textual code, to derive meaning from a minimal number of visual cues. Gibson and Levin (1975) see the reader using context, redundancy, and prior experience with the language in order to derive meaning economically and efficiently. Goodman's (1968) "guessing game" model portrays the reader as using available prior information to hypothesize the meaning of the

upcoming text, providing a basis for scanning the text for cues. These models support the notion that if sentence patterns occur in varying frequencies, the presence of more common or familiar patterns would facilitate the ease with which the text could be read. Strickland (1962) developed a basis for measuring sentence-pattern familiarity in terms of children's oral language. Bormuth (1964) found a non-significant association between this variable and passage cloze scores, but did find their relationship to be curvilinear.

The present study was conducted in order to identify which patterns recur in student writing and in widely-used text, and to determine whether patterns which recur more often, if present, are associated with lower reading difficulty.

Methods

The study was conducted in two phases. In the first phase, samples of student writing and textbook prose were scanned to identify those syntactic patterns which occurred and to count how often each recurred. To do this, the words on the Basic Elementary Reading Vocabularies (Harris and Jacobson, 1972) were classified according to their part-of-speech function using the scheme developed by Fries (1952). Next, ten samples of writing and text, one each for writing and text in grades 2, 4, 6, 9, and 12, were divided into t-units (Hunt, 1964) in order to provide standard sentence-marking throughout the samples. These t-units were scanned using computer processing procedures developed at the University of Virginia; the procedures scanned each t-unit, looked up each word in the classified vocabulary, assigned a coded

character for each word to indicate its syntactic classification, and catalogued a string of these codes for the words in the t-unit. A series of FORTRAN programs was used to count how often the strings occurred in each of the ten samples of prose. The final product of this phrase was a master listing of syntactic patterns and their recurrence in the ten text and writing samples.

In the second phase, this master listing was used to provide an index of the familiarity of sentence patterns in a set of 161 passages from the McCall-Crabbs Standard Test Lessons in Reading (McCall and Crabbs, 1961). The passages were marked off into t-units, and the t-units were converted into coded syntactic strings as was done for the text samples. Each string was located in the master listing to find its recurrence in the ten text samples. These ten values were averaged for the strings in the passage to provide ten values for the passages as a whole. These ten indices were later analyzed in relation to the readability of the passage. Three counts were made of the number of t-units in each passage which were one to ten words in length, eleven to twenty words in length, and twenty-one to thirty words in length. This provided a rich measure of t-unit length which could be controlled in the analysis. Finally, the criterion measures of passage comprehensibility were recorded. The McCall-Crabbs seventy percent criterion (MC70) was used; this is the mean grade equivalent of pupils who answered seventy percent of the comprehension questions on the passage correctly in McCall-Crabbs norming studies.

Analysis

Multiple regression analyses were calculated using the McCall-Crabbs seventy percent criterion as the dependent variable.

In the first analysis, the restricted equation was based on estimates of readability provided by the Harris-Jacobson formula (based on sentence length, word length, word-list frequency, and presence of sound-letter irregularities) and the three counts of t-unit length. The full model added the ten indices of pattern commonness. The following multiple R values were obtained:

TABLE 1

Variables	Multiple R	R^2	F-ratio	df	P
Pattern length counts H-J readability estimates	.62169	.38650	---	---	
Pattern commonness variables	.65055	.42321	1.040	9,147	> .05

The table shows that the pattern commonness variables increase the overall strength of the estimation somewhat, but not to a significant degree.

Since Bormuth's (1964) study had indicated a curvilinear association, the pattern-commonness variables were squared and the analysis was made again:

TABLE 2

Variables	Multiple R	R^2	F-ratio	df	P
Pattern length counts H-J readability estimates	.62169	.38650	---	---	---
Pattern commonness variables and squares	.67664	.45784	2.15	9,147	< .05

With the use of squared terms to account for a curvilinear relationship between passage difficulty and sentence-pattern familiarity, a significant increment in the estimation is obtained

Discussion

The association between the commonness of sentence patterns in a passage of text and the readability of the passage was found to be significant and curvilinear. Subsequent analysis of the data have been performed, using as the criterion the comprehensibility of subsets of the passages to students in grades 4, 6, 9, and 12. These analyses suggest that the usefulness of sentence-pattern familiarity is strongest for older readers. Apparently, use of these frequency-based expectancies depend on the level of difficulty of the passage, and on the ability of the reader.

Although very preliminary in nature, it can be concluded that frequency-based expectancies of syntax are useful to maturing readers in certain textual situations.

Conclusions

This suggests that estimates of readability used to construct a sequence of materials which are graded in difficulty should account for the commonness of syntactic patterns. It also suggests that the use of pattern familiarity, or the ability to contend with unfamiliar patterns, are reading skills which should be developed in instructional programs. This could be accomplished through the exposure of learning readers to a large, diverse amount of natural text to give them a basis for syntactic expectancies. Further, games or exercises which work with these expectancies to predict upcoming text seem to be supported.

REFERENCES

Bormuth, John R. Relationships Between Selected Language Variables and Comprehension Ability and Difficulty. U.S. Office of Education Cooperative Research Project Number 2082, 1964.

_____. Development of Readability Analyses. U.S. Office of Education Cooperative Research Project Number 7-0052, 1969.

Botel, Morton, and Granowski, Alvin. Syntactic Complexity Formula. National Council of Teachers of the English, The Research Instruments Project, 1972.

Dale, Edgar, and Chall, Jeanne. "A formula for predicting readability (and) instructions. Educational Research Bulletin, 27 and 28, 1948.

Dolch, E. Graded reading difficulty. Problems in Reading, Ch. 21, Champaign, Ill.: The Garrard Press, 1948.

Fries, C. C. The Structure of English. New York: Harcourt, Brace, and Company, 1952.

Fry, E. B. A readability formula that saves time. The Journal of Reading, 11, 1968.

Gibson, Eleanor, and Levin, Harry. The Psychology of Reading. Cambridge, Mass.: The MIT Press, 1975.

Goodman, Kenneth. The Psycholinguistic Nature of the Reading Process. Detroit: Wayne State University Press, 1968.

Gray, W. S., and Leary, B. E. What Makes a Book Readable. Chicago: University of Chicago Press, 1935.

Granowski, Alvin. A Formula for the Analysis of Syntactic Complexity of Primary Grade Reading Materials. Unpublished doctoral dissertation, University of Pennsylvania, 1971.

Harris, Albert, and Jacobson, M. D. Basic Elementary Reading Vocabularies. New York: Macmillan and Company, 1972.

_____, and Jacobson, M. D. Revised Harris-Jacobson readability formulas. College Reading Association, Bethesda, Maryland, 1974.

Hunt, Kellogg. Differences in Grammatical Structure Written at Three Grade Levels. U.S. Office of Education Cooperative Research Project Number 1998, 1964.

Lively, B. A., and Pressey, S. L. A method for measuring the "vocabulary burden" of textbooks. Educational Administration and Supervision, 9, 1923.

McCall, and Crabbs. Standard Test Lessons in Reading. New York: Teachers College Press, 1961.

Smith, Frank. Understanding Reading. New York: Holt, Rinehart, and Winston, 1971.

Spache, George. A new readability formula for primary grade reading materials. Elementary School Journal, 53, 1953.

Strickland, R. The language of elementary school children: its relationship to the language of the reading textbooks and the quality of reading in selected children. Bulletin of the School of Education, 38, 1962.

Vogel M., and Washburne, C. An objective method of determining grade placement of children's reading material. Elementary School Journal, 28, 1928.